

Human and Artificial Intelligence for the Society of the Future
European Distance and E-Learning Network (EDEN) Proceedings
2020 Annual Conference | Timisoara, 22-24 June, 2020
ISSN 2707-2819
doi: 10.38069/edenconf-2020-ac0022



WIKIPEDIA, A SOCIOTECHNICAL RESOURCE?

Teresa Cardoso, Filomena Pestana, Universidade Aberta, Laboratório de Educação a Distância e Elearning (LE@D), Portugal

Abstract

The process of digitization is transversal to the different domains of society, including the digitization of encyclopaedias, namely Britannica Online (surpassing its own centennial print edition). In turn, it was surpassed by Wikipedia, which emerged from the advent of web 2.0. Due to the cyclopean proportion that it reached throughout its 19 years of existence, it fulfils an old dream, i.e. to gather, in one place, all the knowledge of humanity – thus, somehow constituting the new library of Alexandria. It also responds to the demands of the digital world, namely by having followed a path aligned with digital and technological evolution, such as the constitution of a community that integrates both human and non-human agents. As Wikipedia is an essential tool when searching on the internet, our aim is to contribute to the understanding of this open resource. Moreover, we address the question of Wikipedia being a sociotechnical system, where bots, non-human collaborators, intervene, either in repetitive automated editions, or progressively expanding its scope, in a sophistication of actions that are not restricted to the content of the articles, but, instead, expand to socialize with the community of users.

Introduction

We live in a globalized society, with socio-cultural implications that incorporate the so-called new technologies, more commonly known as ICT – information and communication technologies. ICT are the backbone of the networked society, which Castells (2003) perceives as a new social morphology; he also mentions that the diffusion of networks significantly modifies the action and the results of the processes of experience, power and culture (at the core of which we can find encyclopaedias, particularly Wikipedia, as we develop below).

This background evokes to us, among others, the concept of cyber-culture, a cultural and social phenomenon defined by Lévy (1999) as a set of techniques that integrate both material and intellectual techniques, to which are added practices, attitudes, ways of thinking and values that interact with each other. For him, it is through the connection

that virtual communities are built – communities based on affinities, interests and common goals that give rise to the collective intelligence. Thus, he defines it as a globally distributed intelligence, which is constantly valued in real time and which leads to the mobilization of skills (Lévy, 1997).

Moreover, it is by understanding the network as an educational interface that integrates and provides the opening and sharing of knowledge, according to Cardoso, Pestana, and Brás (2018), that we will reflect on the current encyclopaedia – Wikipedia. We'll start by presenting and characterizing this online digital phenomenon, and then focus on the role of human and non-human agents so as to sustain that it is a sociotechnical resource.

Wikipedia, an online digital encyclopaedia

Thacz (2015) argues that encyclopaedias assume themselves as historical artifacts that have offered, over time, an understanding of the world at the time they are written, since they provide a body of knowledge at a given moment. It is in this context that he highlights the relevance of understanding today's encyclopaedia, Wikipedia, namely to identify what constitutes our encyclopaedia of knowledge and what are its limits, how it is organized with regard to the knowledge taxonomy, the creation and edition of articles, the distribution of tasks, the hierarchies of those who contribute to its construction – in short, how the whole Wikipedia project works. We follow this reasoning, and so we are interested in understanding not only the products of Wikipedia (its articles, i.e. its most visible elements), but also the processes involved, specially the role that bots assume in this sociotechnical system.

Looking now more in-depth into Wikipedia, it has nearly two decades, as it was officially launched by Jimmy Wales and Larry Sanger on January 15th 2001. This online encyclopaedia called Wikipedia appeared first in English, then in other languages. The Portuguese edition was made available in the same year in June. One way to understand the Wikipedia project is through the interpretation of the narrative that is given to us in its access page. Thus, the Uniform Resource Locator (URL) www.wikipedia.org directs users to Wikipedia, where it is possible to see an incomplete spherical puzzle in the centre, and in which each piece includes the letter “w” in different alphabets. The spherical shape is not complete and is assumed as the current Wikipedia logo. It should be noted that this was not always the logo used (cf. Pestana, 2014). Leitch (2014) associates the logo with the cooperative and global nature of the project, showing itself as an unfinished work. Around it you can find a set of access portals that Wikipedia assumes as relevant. In the words of Tankersley (2016), they are the “top ten viewed wiki's by language”. The languages represented are English, French, German, Italian, Polish, Spanish, Russian, Japanese, Portuguese and Chinese. Each has either a reference to the updated number of existing

articles or a translation of motto “The Free Encyclopedia”. The narrative that is transmitted to us through the homepage allows us to highlight the magnitude of the project and the institution that supports it, i.e. the Wikimedia Foundation.

We emphasize that the idea of grouping knowledge in a single place has been a dream that dates back to ancient Greece. These attempts, in a more recent period, have also been present. Because understanding its origin also involves integrating the Wikipedia project into the set of projects that somehow preceded or are contemporary to it, as well as its particular characteristics. Thus, according to Ayers, Matthews, and Yates (2008), Lih (2009), Reagle (2010) and Rosenzweig (2006), Wikipedia was Jimmy Wales’ second attempt to build a free online encyclopaedia. Nupedia’s first designated attempt started in March 2000 and was a failure; also, like Wikipedia, it had as main actors Jimmy Wales and Larry Sanger. Its objective was to design articles that were developed in a top/down structure, characterized by a heavy and academic model, based on a peer-review process (Ayers, Matthews, & Yates, 2008; Rosenzweig, 2006). In other words, it was still organized like traditional encyclopaedias, with an expert advisory board and a long review process. Nupedia can be considered as the genesis of Wikipedia (for a more detailed description see Pestana, 2014). Wikipedia has had a huge growth since its beginning, presenting, according to Wales (2012; 4m: 21s), “a very bright future”. In fact, as he recognises, “Wikipedia has become part of our infrastructure and life” (ibid.; 2m: 24s).

However, a distinctive aspect, in comparison to other encyclopaedias, is the fact that the articles are not necessarily written by experts in the field, nor their review is necessarily done by peers. Moreover, Wikipedia has presented, in its first decade of existence, an evolution in the way in which articles are published (Rosenzweig, 2006). Although maintaining the same principles, there is a set of Fundamentals, Rules and Recommendations that support their publication. According to Ayers, Matthews, and Yates (2008) and Lih (2009), they are the pillars that have evolved since the three founding principles, established by Larry Sanger – 1. Neutral Point of View (NPOV), 2. Verifiability (V), and 3. No Original Research (NOR) –, and are presented as the Fundamentals of all Rules and Recommendations. Those Pillars are the following: Encyclopaedism; Neutrality of point of view; Free license; Community coexistence; Freedom in the rules.

Another important feature is the fact that although Wikipedia is a phenomenon emerging within the Web 2.0, it has been evolving and can be integrated within the phenomenon of Web 3.0 too. For example, the availability of the Objective Revision Evaluation Service (ORES) is:

“a new artificial intelligence service designed to improve the way editors maintain the quality of Wikipedia. This service empowers Wikipedia editors

by helping them discover damaging edits and can be used to immediately ‘score’ the quality of any Wikipedia article. We’ve made this artificial intelligence available as an open web service that anyone can use.”
(Halfaker & Halfaker, 2015)

As also pointed out, for example, by Clément and Guitton (2015) or Niederer and Dijck (2010), Wikipedia involves human agents and non-human agents, which makes it a sociotechnical resource, as we further develop in the following section.

Wikipedia, a sociotechnical resource

In this new era of Wikipedia, one of its major sister projects is Wikidata. In fact, Leitch (2014; p.120) refers that “The Wikimedia Foundation has taken a step into the world of Web 3.0 with Wikidata [...] [it] is designed to provide a more centralized and highly structured repository of information for all the languages used by members of the Wikipedia community and their computers”. According to Manske (2014), Wikidata is “a free knowledge base that can be read and edited by humans and machines alike”.

Focusing again on Wikipedia, with regard to human agents, Wikipedia is built with the collaboration of a huge and diverse group of volunteers, who produce the articles – the wikipedians. They may have different statutes, according to certain requirements needed to apply for the different roles; its hierarchy is presented below. For Niederer and van Dijck (2010; p.6), the dynamics generated in the creation of content between human and non-human agents are a crucial aspect of Wikipedia’s performance, i.e. “The online encyclopaedia’s success [...] [is] based on sociotechnical protocological control, a combination of its technical infrastructure and the collective ‘wisdom’ of its contributors”. And, as seen on Figure 1, the permission level depends not only on the type of profile of wikipedians (from blocked users to developers, including human and non-human agents), but also on their functions or responsibilities (four are specified, including the management of Wikipedia access accounts).

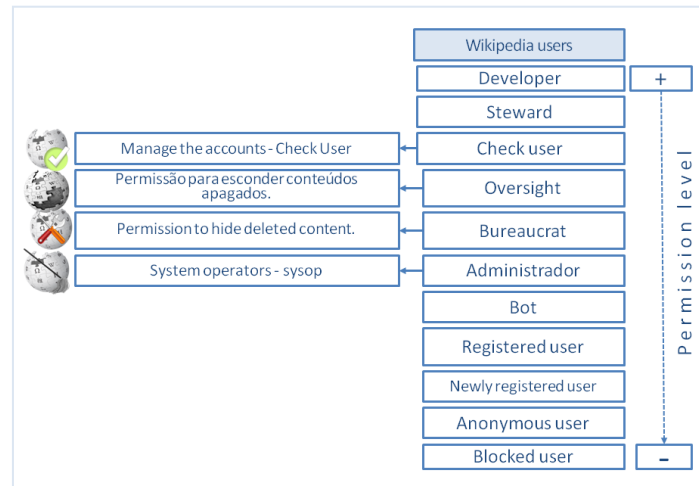


Figure 1. Hierarchy of human and non-human agents in the Wikipedia project (from Niederer and van Dijck, 2010).

As represented on figure 1, the developer and the steward (on top) are associated not only with the Wikipedia project, but also with other Wikimedia Foundation projects. The former, at the software level, and the latter, related to the several Wikipedia sister projects. In the hierarchy above, we also highlight the existence of bots, since they help to control (automatically) the edition of the Wikipedia articles. Bots are seen as non-human collaborators who work under the supervision of humans with regard to repetitive and successive editions (d'Andréa, 2011). According to Niederer and van Dijck (2010), the Wikipedia community, due to the enormous amount of data generated, was forced to implement bots with a view to increasing the quality and structure of the data. Zheng, Albano, Vora, Mai, and Nickerson (2019: 4) further acknowledge that:

“The first Wikipedia bot appeared in October 2002 [...] Currently, Wikipedia uses a ‘decentralized, consensus-based model’ to regulate bot-related work. [...] Contributors who want to develop and deploy a bot are expected to submit a bot approval request that provides information about the bots’ functions, the bots’ programming language, and the estimated number of pages affected. Then a Bot Approvals Group (BAG) run by experienced and trusted developers will go over the request and discuss its potential influence.”

Clément and Guitton (2015) find that cohabitation between human and non-human agents is a consequence of the drastic increase in social networks and virtual communities, which reduces the individual response capacity of human agents in favour of emerging tools that help them (i.e. bots). Moreover, they identify two opposite types of bots: the servant bots, with a low level of autonomy, which facilitate the work of users working behind the scenes; the policing bots, controlled by the administrators, act proactively with a high level of autonomy, restricting the work of users and enforcing Wikipedia’s guidelines and

standards. For the authors, the bots were analysed from the following characteristics: the nature of the owner (administrator and registered user); the function of the bot (protection from vandalism, patrolling of materials with regard to copyright, and notices to users) or job assistance (replacing templates, correcting or redirecting false links); the places where changes were made (articles, user page, discussion pages), or behind the scenes (categories, portals and help); automatic way of acting, that is, in “opt-out way” or “opt-in”; degree of autonomy of the bot that is embodied at a high level (some degree of evaluation of user contributions) and at a low level (purely factual or issue management, such as correction of false links or distribution of e-mail around users of the Wikipedia).

Vandalism, “the malicious modification or editing of articles, is a serious problem for free and open access online encyclopedias such as Wikipedia” (Tran, 2015; p.ix). According to Hamiti, Susuri, and Dika (2015), it was only in the beginning of 2006 that bots were programmed to reverse the vandalism, a reversal that is made after the automatic scanning of the editions. It should be noted that previously this work was carried out manually, since vandalism has existed since the beginning of Wikipedia. In their research, the authors conclude that the greatest advances have been achieved by the English, German, French and Spanish communities, with a residual impact on the others. Tran (2015; p.14) identifies two main categories in the anti-vandal software: “automatic detection (bots) and assisting users (editing applications)”; each category includes different editing applications (see below, Table 1 and Table 2).

Table 1: Bots to combat vandalism (from Tran, 2015; p.14)

Notable Bots	Prevailing editing applications
Anti-Vandal Tool	“a bot that monitors the feed of all edits on Wikipedia as they occur. Vandalism is detected by matching words in the edit to a list of vandal words used in past vandalism cases”
ClueBot	“the most active counter-vandal bot from 2007 to 2011. When this bot inspects an edit, a score is determined from a variety of pattern matching heuristics that includes large changes, mass deletes, controversial topics, targeted celebrities, incorrect redirects, vulgar words, minor sneaky changes and others that are added as certain types of vandalism are discovered.”
ClueBot NG35	“the successor to ClueBot and also the first Wikipedia counter-vandalism bot to use machine learning algorithms to improve detection rate and lower false positives.”

Table 2: Most relevant editing applications to combat vandalism (from Tran, 2015; p.16)

Notable editing applications	Prevailing editing applications
Huggle	"a browser application that allows fast viewing of in-coming edits. It allows users to identify vandalism or non-constructive edits, and to quickly revert them."
STiki	"a cross-platform application for trusted users to detect and revert vandalism and other non-constructive edits."
Snuggle	"a browser application designed to allow experienced editors to observe the activities of new editors and distinguish vandals and nonvandals. This application was developed from research [Halfaker et al., 2014] to address the decline in retention of new Wikipedia users. The interface provides four categories to classify edits analogous to STiki, but allows viewing of an editor's editing history and personal messaging to provide feedback to (new) users."

However, and even though bots, as Wikipedia non-human agents, tend to assist human agents in the different work areas of this encyclopaedia (Benevolent bots), according to Tsvetkova, García-Gavilanes, Floridi, and Yasseri (2017), there are also bots outside this sociotechnical system that promote a set of malevolent actions (Malevolent bots). In addition to this situation they identify that sometimes there are bots within the system that conflict. But, Geiger and Halfaker (2018; p.5), in the context of bot governance, state that the situation is already abundantly debated and argue that:

"Wikipedia's model of automation regulation is generally based on the same principles as editing Wikipedia articles: decentralized consensus-building, scaffolded by formalized policies and processes. Just like with editing articles, Wikipedians sometimes get into intense debates, conflicts, and controversies about whether the Bot Approvals Group ought to approve or deny a particular bot developer's application — and in some cases, about whether the BAG should rescind an approval for various reasons [...] Unapproved bots are not allowed to edit encyclopedia articles; they must get prior approval from the BAG for a specific, well-defined task. The BAG is a standing committee of bot developers and non-developers tasked with reviewing proposals about new bots in line with the community-authored Bots policy."

Final remarks

It is not difficult to imagine the volume of interactions of a project with the magnitude of Wikipedia, namely if we consider that it is an encyclopaedia in 309 different languages. Actually, "10 Wikipedias [...] were closed and moved to the Wikimedia Incubator for further development, so there is a current total of 299 active Wikipedias" (List of Wikipedias, 2020). Besides, the fact that theoretically everyone can edit caters for editions

that in some cases fit in the vandalism categories, and in other cases may convey errors in content or yet disrespect in some way the rules of wikification. Therefore, the supervision of the fundamentals, pillars and rules by which the community of wikipedians is guided requires enormous vigilance.

As Halfaker, Geiger, Morgan, and Riedl (2012: 683) recognize, “Wikipedia has changed from the encyclopedia that anyone can edit to the encyclopedia that anyone who understands the norms, socializes himself or herself, dodges the impersonal wall of semi-automated rejection, and still wants to voluntarily contribute his or her time and energy can edit”. In fact, over time, bots have come to play an increasingly important role in Wikipedia, an online digital encyclopaedia that, as argued in this text, is a sociotechnical resource. This system that consists of both human and non-human agents. Ultimately, as educators and trainers, advocating for the quality of open resources, we value pedagogical experiences such as those developed within the Wikipedia in Education Program (e.g. Pestana, 2018). We also value research studies on “the dynamics of the bots with respect their changing functions and the effects of such changes on human editors may be important for understanding the dynamic of coordination in knowledge-creating processes. More broadly, this may help us understand the changing ways automation affects knowledge production and human work.” (Zheng, Albano, Vora, Mai, & Nickerson, 2019; p.15)

References

- Ayers, P., Matthews, C., & Yates, B. (2008). *How Wikipedia Works and How Can Be Part of It*. San Francisco, CA: Starch Press.
- Cardoso, T., Pestana, F., & Brás, S. (2018). A Rede como Interface Educativa: uma Reflexão em Torno de Conceitos Fundamentais. *Revista Interfaces Científicas – Educação*, 6(3), 41-52. doi: 10.17564/2316-3828.2018v6n3p41-52
- Castells, M. (2003). *A Sociedade em Rede*. A Era da Informação: Economia, Sociedade e Cultura, v. 1. Lisboa: Fundação Calouste Gulbenkian.
- Clément, M., & Guitton, M. (2015). Interacting with bots online: Users’ reaction to actions of automated programs in Wikipedia. *Computers in Human Behavior*, 50(C), 66-75. doi: 10.1016/j.chb.2015.03.078
- d’Andréa, C. (2011). *Processos editoriais auto-organizados na Wikipédia em português: a edição colaborativa de “Biografias de Pessoas Vivas”* (Doctoral dissertation), Universidade Federal de Minas Gerais. Retrieved from <https://bit.ly/2UspcjB>
- Geiger, R., & Halfaker, A. (2018). Operationalizing Conflict and Cooperation between Automated Software Agents in Wikipedia: A Replication and Expansion of “Even

- Good Bots Fight”. *Proceedings of ACM on Human Computer Interaction*, 1(2), Article 49. CSCW 2018. doi: 10.1145/3134684
- Halfaker, A., & Halfaker, D. (2015, November 30). Artificial intelligence service gives Wikipedians ‘X-ray specs’ to see through bad edits. Diff [Blog post]. Retrieved from <https://bit.ly/2HmCVVK>
- Halfaker, A.; Geiger, R.; Morgan, J., & Riedl, J. (2012). The Rise and Decline of an Open Collaboration System: How Wikipedia’s Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist*, 57(5), pp. 664-688.
<https://doi.org/10.1177/0002764212469365>
- Hamiti, M., Susuri, A., & Dika, A. (2015). Machine Learning and the Detection of Anomalies in Wikipedia. *Proceedings of the 19th International Conference on Circuits Systems Communications and Computers*, 242-246. Retrieved from <https://bit.ly/2TFkvG0>
- Leitch, T. (2014). *Wikipedia U. Knowledge, authority, and liberal education in the digital age*. Maryland: Johns Hopkins University Press.
- Lévy, P. (1997). *A Inteligência Colectiva. Para uma antropologia do ciberespaço*. Lisboa: Instituto Piaget.
- Lévy, P. (1999). *Cibercultura. Editora, 34*.
- Lih, A. (2009). *The Wikipedia Revolution. How a bunch of nobodies created the world’s greatest encyclopedia*. Great Britain: Aurum Press.
- List of Wikipedias. (2020, May 27). Meta, discussion about Wikimedia projects. Retrieved 16:14, May 29, 2020 from https://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias&oldid=20120861
- Manske, M. (2014). Wikidata: A free knowledge base that can be read and edited by humans and machines alike Retrieved from <https://bit.ly/2IXMgpb>
- Niederer, S., & van Dijck, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society*, 12(8), 1368-1387.
<https://doi.org/10.1177/1461444810365297>
- Pestana, F. (2014). *A Wikipédia como recurso educacional aberto: conceções e práticas de estudantes e professores no ensino superior online* (Master dissertation). Universidade Aberta, Lisboa. Retrieved from <https://goo.gl/YbjuzS>
- Pestana, F. (2018). *A Wikipédia como Recurso Educacional Aberto: um contributo para o Programa Wikipédia na Universidade* (PhD Thesis). Universidade Aberta, Lisboa. Retrieved from <https://bit.ly/2DT997r>

- Reagle, J. (2010). *Good Faith Collaboration: The Culture of Wikipedia*. Massachusetts: The MIT Press.
- Rosenzweig, R. (2006). Can History be Open Source? Wikipedia and the Future of the Past. *The Journal of American History*, 93(1), 117-146. doi: 10.2307/4486062
- Tankersley, D. (2016, August 16). Opening the door to a new look: improving Wikipedia.org. Diff [Blog post]. Retrieved from <https://bit.ly/2UsfHku>
- Tran, K. (2015). *Detecting Vandalism on Wikipedia across Multiple Languages* (Doctoral dissertation). The Australian National University. Retrieved from <https://bit.ly/2Houvgn>
- Tsvetkova, M.; García-Gavilanes, R.; Floridi, L. & Yasseri, T. (2017). Even good bots fight: The case of Wikipedia. *PLoS ONE*, 12(2), e0171774. <https://doi.org/10.1371/journal.pone.0171774>
- Wales, J. (2012). Jimmy Wales on the Future of Wikipedia. FOX Business. [Video]. Retrieved from <https://bit.ly/2XLGiuU>
- Zheng, L.; Albano, C.; Vora, N.; Mai, F., & Nickerson, J. (2019). The Roles Bots Play in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction*, Article No.: 215. <https://doi.org/10.1145/3359317>